

The Second Laws for an Information driven Current through a Spin Valve

Philipp Strasberg¹, Gernot Schaller¹, Tobias Brandes¹, and Christopher Jarzynski²

¹ *Institut für Theoretische Physik, Technische Universität Berlin, Hardenbergstr. 36, D-10623 Berlin, Germany*

² *Department of Chemistry and Biochemistry and Institute for Physical Science and Technology, University of Maryland, College Park, Maryland 20742, USA*

We propose a physically realizable Maxwell’s demon device using a spin valve interacting unitarily for a short time with electrons placed on a tape of quantum dots, which is thermodynamically equivalent to the device introduced by Mandal and Jarzynski [PNAS **109**, 11641 (2012)]. The model is exactly solvable and we show that it can be equivalently interpreted as a Brownian ratchet demon. We then consider a measurement based discrete feedback scheme, which produces identical system dynamics, but possesses a different second law inequality. We show that the second law for discrete feedback control can provide a smaller, equal or larger bound on the maximum extractable work as compared to the second law involving the tape of bits. Finally, we derive an effective master equation governing the system evolution for Poisson distributed bits on the tape (or measurement times respectively) and we show that its associated entropy production rate contains the same physical statement as the second law involving the tape of bits.

PACS numbers: 05.70.Ln, 05.60.Gg, 05.40.-a, 05.70.-a, 73.23.Hk

I. INTRODUCTION

Traditionally, thermodynamics is a theory that describes systems exchanging energy and entropy with idealized heat reservoirs and work sources. It provides two fundamental laws every system must obey: the first law (energy balance) and the second law (entropy increase). Reconciling these macroscopic laws with a microscopic picture of atoms and molecules in motion raises subtle issues. As far back as the nineteenth century, Maxwell speculated that an external agent, using microscopic information to control the state of a system, might be able to circumvent the second law [1]. This agent is now known as ‘Maxwell’s demon’. Accepting the possibility of ‘intelligent interventions’ (due to idealized measurement and information processing devices such as computers), it is widely accepted that the correct application of Landauer’s principle [2] will save the second law, as noted by Bennett [3] in his analysis of Szilard’s famous engine [4]. Note however that Maxwell’s demon can also be ‘exorcized’ without reference to intelligent interventions, by showing that any device designed to exploit molecular scale fluctuations to violate the second law, must fail or must produce a large dissipation somewhere else as analysed for specific models by Smulochowski [5] and Feynman [6]. See also Ref. [7] for opposing arguments and Refs. [8, 9] for other illustrative models.

Besides the possibility of actively measuring and controlling the system, Bennett noted that a tape of bits in a low entropy state might act as a thermodynamic resource while the tape randomizes itself [3], which can be regarded as another form of a Maxwell demon in which the low entropy state of the tape allows for the rectification of thermal fluctuations, for instance to lift a mass or cool a cold reservoir. A particular mathematical model was proposed and analyzed by Mandal and Jarzynski [10] and has been followed by further models exploring the possibility of Maxwell’s demon devices incorporating an

information reservoir in the thermodynamic description [11–16].

However, the models proposed above are rather abstract and it would be desirable to find a *physical* model able to reproduce the same results. By ‘physical’ model we mean that we start with a well-defined and well-motivated Hamiltonian describing an explicit, physically realizable system, and, following standard procedures (e.g. as used to derive master equations), we end up with a mathematical description equivalent to those above. In this paper we will indeed show that a quantum dot spin valve with perfectly polarized leads allowing only for one sort of spins to tunnel through, which interacts with a tape of electrons causing spin flips, provides a physical implementation of the device proposed in Ref. [10] in the sense that it has identical dynamics and thermodynamics (though a quite different physical interpretation). A similar model was already put forward by Datta in Ref. [17] where the ‘impurities’ in his model correspond to the electrons on the tape in our model. For another physical realization also see Ref. [18].

The models discussed above all rely on measurements or interactions with a bit at predetermined discrete times or intervals. The resulting dynamics can in general not be formulated as a differential equation anymore. An alternative approach to feedback control relies on a continuous measurement scheme and the fact that a master equation (ME) can be unraveled in terms of trajectories representing the actual state of the system. This approach yields an effective differential equation for the system dynamics and has been extensively used in the field of open quantum systems and quantum information [19]. Furthermore, it was successfully applied to construct Maxwell demon like feedbacks and to study the thermodynamics of such systems [20–22]. It was shown that such feedback schemes can also be formulated within an inclusive approach not relying on any phenomenological measurements or feedback actions [9]. In addition

to the ME picture, the thermodynamic implications of continuous feedback schemes were studied for Langevin dynamics as well [23, 24].

Aside from the quest to design illustrative devices, which are able to rectify thermal fluctuations by some sort of information processing, the question naturally arises as to whether it is possible to treat the different approaches above in a unified framework [25–29]. To address this question with our model – following the three cases investigated by Barato and Seifert [26] – we will design a measurement-based feedback scheme, which has identical system dynamics as the tape model, and we will derive an effective ME for Poisson distributed bits on the tape or measurement times respectively. We will compare the different forms of the second laws and we will show that the second law for discrete feedback control can provide a smaller, equal or larger bound on the amount of extractable work as compared to the second law involving the tape of bits. For the effective ME we will show that its associated entropy production represents the same physical statement as the second law involving the tape of bits.

Outline: We start with the description of the model including the system, the tape of bits and the system-bit interaction in Sec. II and we show that it is equivalent to a Brownian ratchet demon. Sec. III is then devoted to a thorough study of its thermodynamics. In Sec. IV we will study a measurement based feedback scheme, which yields identical system dynamics but different thermodynamics and we will compare the two different second laws. Finally, we derive an effective ME for a Poisson distributed tape in Sec. V and discuss its thermodynamic behaviour in relation to the other approaches. In Sec. VI we discuss our results.

II. PHYSICAL IMPLEMENTATION OF THE MODEL

A. System Description

The system we want to control – either by a tape of bits or by feedback control (see Sec. IV) – consists of a quantum dot in the ultra strong Coulomb blockade regime attached to two spin-polarized ferromagnetic leads, which preferably allow one spin direction to tunnel. Such quantum dot spin valves can be experimentally realized by molecular quantum dots [30, 31]. The quantum dot is either empty or filled by an electron with spin up or spin down, thus the system Hilbert space is three dimensional, $\mathcal{H}_S = \text{span}\{|0\rangle, |\uparrow\rangle, |\downarrow\rangle\}$, and the Hamiltonian of the system is assumed to be $H_S = \epsilon_s(|\uparrow\rangle\langle\uparrow| + |\downarrow\rangle\langle\downarrow|)$. For perfectly and oppositely polarized leads the usual Born-Markov secular ME predicts that the time evolution of the probability vector $\rho_S = (p_0, p_\uparrow, p_\downarrow)$ is given by the rate equation $\partial_t \rho_S(t) = \mathcal{W} \rho_S(t)$ where the Liouvillian \mathcal{W}

can be split into two parts $\mathcal{W} = \mathcal{W}_L + \mathcal{W}_R$:

$$\begin{aligned} \mathcal{W}_L &= \Gamma \begin{pmatrix} -f_L & 1-f_L & 0 \\ f_L & -(1-f_L) & 0 \\ 0 & 0 & 0 \end{pmatrix}, \\ \mathcal{W}_R &= \Gamma \begin{pmatrix} -f_R & 0 & 1-f_R \\ 0 & 0 & 0 \\ f_R & 0 & -(1-f_R) \end{pmatrix}. \end{aligned} \quad (1)$$

A detailed derivation of this equation even for arbitrary polarized leads is given in Ref. [32] following the steps stated in [46]. Furthermore, within the framework of the Born-Markov secular ME, we do not have to include the off-diagonal elements of the system density matrix (“coherences”) because in the energy eigenbasis they do not affect the populations but simply follow damped oscillations and die out in the long-time limit. The transition rates of the Liouvillian \mathcal{W} are specified by a global tunneling rate $\Gamma > 0$ and Fermi functions $f_\nu = f_\nu(\epsilon_s) = [e^{\beta_\nu(\epsilon_s - \mu_\nu)} + 1]^{-1}$, $\nu \in \{L, R\}$, where β_ν and μ_ν are the inverse temperature and chemical potential of lead ν . For the rest of the text we fix the chemical potentials at $\mu_L \equiv \epsilon_s + \frac{V}{2}$ and $\mu_R \equiv \epsilon_s - \frac{V}{2}$ where $V = \mu_L - \mu_R$ denotes the bias voltage. Assuming equal lead temperatures, $\beta_L = \beta_R \equiv \beta$, the Fermi functions fulfill the local detailed balance condition

$$\frac{f_L}{1-f_L} = e^{\beta V/2}, \quad \frac{f_R}{1-f_R} = e^{-\beta V/2} \quad (2)$$

and the steady state $\bar{\rho}_S$ of the system is given by

$$\bar{p}_\uparrow = \frac{e^{\beta V}}{Z}, \quad \bar{p}_0 = \frac{e^{\beta V/2}}{Z}, \quad \bar{p}_\downarrow = \frac{1}{Z} \quad (3)$$

with $Z = e^{\beta V} + e^{\beta V/2} + 1$. It is easy to verify that the electric current through the system vanishes: due to the opposite polarization of the leads the transport through the spin valve is blocked. We note however that this is an idealized description. In practice the leads are not perfectly polarized and phonons in the surrounding medium can induce spin flips in the quantum dot. For simplicity we will neglect these effects because they do not change the general point of our discussion as long as they are relatively weak. For the rest of this paper we will set $\beta \equiv 1$.

B. Tape of Bits

We now describe the tape of bits, which moves frictionlessly past the quantum dot spin valve with constant speed. In contrast to the models of Refs. [10, 11, 13, 15] we assume that each bit interacts with the system only for a short time $\delta t \ll \Gamma^{-1}$ and that the time τ between two successive bits is relatively long, $\tau \gg \delta t$. During this time τ the system evolves according to the Liouvillian \mathcal{W} .

Physically, each bit on the tape represents another quantum dot containing one excess electron, which is

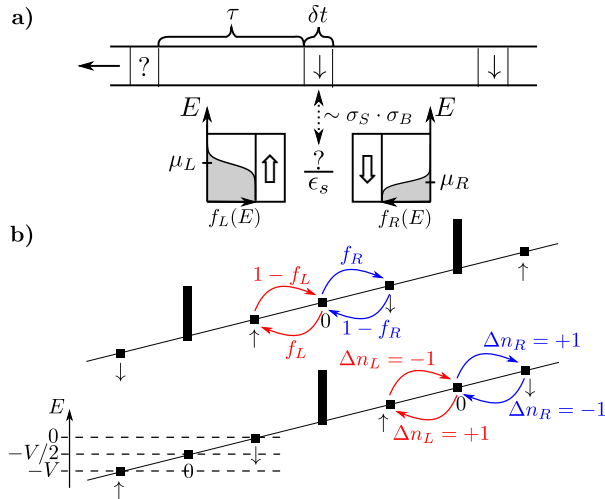


FIG. 1: a) Simple sketch of the system showing the tape of bits coupled to the quantum dot spin valve. b) Equivalent model of the Brownian ratchet showing the two different modes, which can be regarded as shifted versions of an infinite lattice. Transitions between the two different modes are due to the interaction with a bit or due to the feedback.

either in state spin up or spin down. To quantify the portion of spin up and spin down electrons we introduce the excess parameter δ^- , which is defined by

$$\delta^- \equiv p_{b=\uparrow}^- - p_{b=\downarrow}^- . \quad (4)$$

Here, $p_{b=\sigma}^-$ denotes the probability that an *incoming* bit is in state $\sigma \in \{\uparrow, \downarrow\}$. Analogously, $\delta^+ \equiv p_{b=\uparrow}^+ - p_{b=\downarrow}^+$ quantifies the excess of spin up electrons in the *outgoing* tape. As with the quantum dot spin valve, we assume that the spin up and spin down states on the bit dots are energetically degenerate, i.e., $H_B = \epsilon_b(|\uparrow\rangle\langle\uparrow| + |\downarrow\rangle\langle\downarrow|)$. A visualization of the system and bit string is provided in Fig. 1 (a).

C. System-Bit Interaction

Finally, we must specify how the system and the bit interact during the short time δt . We will model the interaction between a filled quantum dot and the n 'th bit by an effective Heisenberg Hamiltonian of the form $J_n(t)\sigma_S \cdot \sigma_{B_n}$. Here $\sigma = (\sigma^x, \sigma^y, \sigma^z)^T$ is the vector of Pauli matrices and $J_n(t) \geq 0$ is a coupling strength whose time dependence is induced by proximity between the system and bit. If the interaction occurs around the time $t = t_n$, we demand that

$$\int_{-\infty}^{\infty} ds J_n(s) \approx \int_{t_n - \delta t/2}^{t_n + \delta t/2} ds J_n(s) = \frac{\pi}{4}, \quad (5)$$

i.e., $J_n(t)$ is approximately zero for $t \notin [t_n - \delta t/2, t_n + \delta t/2]$. Our assumption $\delta t \ll \Gamma^{-1}$ allows us to neglect any

dissipative dynamics arising from the interaction with the leads. The time evolution of the combined system and bit – given that the spin valve quantum dot is initially filled with an electron – is governed by the unitary operator

$$\exp \left[-i \int_{t_n - \delta t/2}^{t_n + \delta t/2} ds J_n(s) \sigma_S \cdot \sigma_{B_n} \right] \sim U_{\text{swap}}. \quad (6)$$

Here, U_{swap} is the unitary swap operator acting on two spins as $U_{\text{swap}}|\sigma_1\sigma_2\rangle = |\sigma_2\sigma_1\rangle$. Thus, if the spins of the system and the bit are the same then nothing happens, but if they are opposite then they both flip. If the quantum dot of the spin valve is initially empty, $|0\rangle$, we assume that the state of the bit does not change during the interval $t_n - \delta/2 < t < t_n + \delta/2$.

The exact form of the time dependent coupling strength $J_n(t)$ is unimportant: as long as $\delta t \ll \Gamma^{-1}$, it only needs to fulfill Eq. (5). In practice, of course, $J_n(t)$ will never exactly fulfill Eq. (5) but will differ by a small amount ϵ such that the true unitary operation we implement is given by $U_{\text{swap}} + \epsilon U_{\text{cor}} + \mathcal{O}(\epsilon^2)$ where the correction term U_{cor} gives rise to quantum coherences between the spin up and down states and a finite probability proportional to ϵ^2 that the swap operation does not succeed. Because our treatment is in any case idealized (for instance, due to the assumed perfect polarization of the leads and the absence of phonons) we will also assume that $\epsilon = 0$, but a more sophisticated treatment would include a discussion about the so called *fidelity* of our *quantum gate* (i.e., our swap operation).

The effective Heisenberg Hamiltonian above can be physically more rigorously justified by considering a Hubbard Hamiltonian for the two quantum dots (spin valve plus bit), introducing a weak electron tunneling term between the dots and for which one can show that the effective low energy interaction is governed by the Heisenberg Hamiltonian above [33]. The fact that this allows the realization of a swap gate is also used for purposes of quantum information processing in quantum dots [34]. In these setups the time dependent coupling $J_n(t)$ is induced by an electrostatic barrier, which can be controlled by a gate voltage, between two spatially fixed neighbouring quantum dots. Thus, we do not necessarily need a tape of bits moving past the spin valve but we could also use a single electron transistor or any other single electron source adjacent to the spin valve and control the interaction by a gate voltage. However, for similarity with the models proposed in Refs. [10–16] we will stick to the picture of a tape of quantum dots pulled along the spin valve. Finally, note that a SWAP operation between two spin qubits in quantum dots was experimentally realized already in 2005 [35].

D. System Dynamics

We are now ready to formulate the dynamical evolution of the system over one cycle of duration τ . For

conciseness we henceforth neglect the small time window δt and simply assume that the swap operation happens instantaneously. We start a cycle with the system bit interaction, eventually leading to a swap operation, and end a cycle shortly before the next bit comes to interact with the system. Once we have reached a periodic steady state we can drop the index n specifying the bit of the tape because the statistical behaviour of the system is the same from one interaction interval to the next. For the interval during which there is no system bit interaction, the time evolution of the system is generated by the Liouvillian (1), i.e.,

$$\rho_S^-(t + \tau) = e^{\mathcal{W}\tau} \rho_S^+(t), \quad (7)$$

where the superscript $- (+)$ denotes the state of the system before (after) the swap operation. The system state after the swap operation is related to the system state before via the transformation

$$\rho_S^+ = \text{tr}_B[U \rho_S^- \otimes \rho_B^- U^\dagger] = \begin{pmatrix} p_0^- \\ p_1^- \frac{1+\delta^-}{2} \\ p_1^- \frac{1-\delta^-}{2} \end{pmatrix} \quad (8)$$

where $U \equiv 1_B \otimes |0\rangle_S \langle 0| + U_{\text{swap}} |1\rangle_S \langle 1|$ and $|1\rangle_S \langle 1| = |\uparrow\rangle_S \langle \uparrow| + |\downarrow\rangle_S \langle \downarrow|$ denotes the projector onto the 1-electron subspace in the spin valve. We have introduced the notation $p_1^- \equiv p_{\uparrow}^- + p_{\downarrow}^-$ and have used the fact that the initial state of the bit is given by $\rho_B^- = p_{b=\uparrow}^- |\uparrow\rangle_B \langle \uparrow| + p_{b=\downarrow}^- |\downarrow\rangle_B \langle \downarrow|$ with $p_{b=\uparrow}^- = (1 + \delta^-)/2$ and $p_{b=\downarrow}^- = (1 - \delta^-)/2$.

Eq. (7) together with Eq. (8) fully specify the time evolution of the *system*, which can now be solved for its periodic steady state. Because this expression is a bit lengthy we will not display it here. Furthermore, we will henceforth measure τ in units of Γ^{-1} , i.e., we can set without loss of generality $\Gamma \equiv 1$.

E. Mapping to a Brownian Ratchet

Although our model is taken from the field of electron transport, we will now show that it can indeed be mapped to a Brownian ratchet. This makes it easier to relate our model to other work in the field of information thermodynamics and Maxwell's demon where many models are based on classical Brownian dynamics, see for instance Ref. [12], which describes a very similar device and Ref. [36] for an experimental realization.

To see why this can be done, let us imagine that we have a particle hopping between three discrete states labeled $\uparrow, 0, \downarrow$ with corresponding energies $E_{\uparrow} = -V, E_0 = -\frac{V}{2}, E_{\downarrow} = 0$. If this particle is immersed in a heat bath, its thermally activated transitions $\uparrow \leftrightarrow 0$ and $0 \leftrightarrow \downarrow$ (note that the direct transition $\uparrow \leftrightarrow \downarrow$ is forbidden) obey local detailed balance exactly as in Eq. (2). This picture is also supported by the fact that the stationary state of the spin valve, Eq. (3), equals a canonical equilibrium state with respect to the energies defined above.

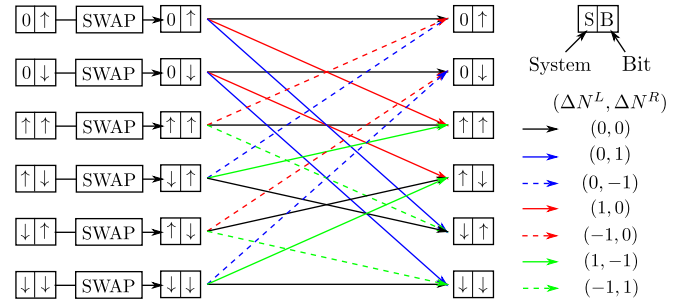


FIG. 2: Sketch of the possible transitions during a single cycle, starting from the left with the initial state $\rho_S^- \otimes \rho_B^-$, followed by a swap operation, and ending after a time τ in the final state, which becomes the new initial state for the next cycle if we replace it with a new bit. The arrows indicate possible transitions and the color code symbolizes the net number of exchanged particles with the left and right reservoir contributing to Eqs. (10) and (11).

The analogy to a Brownian ratchet demon, in which a particle is transported against an external force (or *load*) by switching between two different potentials, is completed by recognizing that our fictitious particle could live on a discrete lattice $(\dots, \uparrow, 0, \downarrow, \uparrow, 0, \downarrow, \dots)$ with impenetrable walls between the states \downarrow and \uparrow , but is allowed to change between two different modes due to the interaction with a bit (or later due to the feedback, Sec. IV). This is summarized in Fig. 1 (b). It is easy to show that the work ΔW against the external load over one cycle is given by $\Delta W = -\frac{V}{2} \Delta N^L + \frac{V}{2} \Delta N^R = -V \Delta N^L$ where ΔN^ν denotes the average number of tunneled particles from lead ν during one cycle and we have used particle number conservation $\Delta N^R = -\Delta N^L$. We will define ΔN^ν more thoroughly in Sec. III A.

III. THERMODYNAMICS

Before we start with the thermodynamic analysis we make two important remarks: first, the full behaviour of the model depends only on three parameter: the voltage $V \in (-\infty, \infty)$, which favors a particular spin direction in the spin valve, the excess parameter $\delta^- \in [-1, 1]$ of the incoming bits and the interval duration $\tau \in (0, \infty)$. Second, our system possesses a symmetry: mapping the parameters (V, δ^-) to $(-V, -\delta^-)$ is physically equivalent to exchanging the labels L and R and \uparrow and \downarrow of the spin valve. When we later plot the phase diagram of our model over $V \times \delta^- = (-\infty, \infty) \times [-1, 1]$ this symmetry allows us to limit the discussion to one half of the phase space because the behaviour in the other half follows by symmetry arguments.

A. Work, Heat and Bit Statistics

In appendix A we show that the work expenditure ΔW_{pull} to pull the tape in our idealized setup (assuming a frictionless sliding tape) is truly zero. Physically, this is due to the two facts that we treat the swap operation unitarily and that the spin up and down states are energetically degenerate in the system and bit quantum dot.

Next, we determine the flow of electrons at the left or right contact, defined to be positive if the electron hops out of the reservoir into the system. This is most easily done by looking at all possible transitions during one cycle and associating the corresponding change in particle number with it as shown in Fig. 2. For instance, for the number of tunneled particles at the left contact we must add the terms (following the diagram in Fig. 2 from top to bottom)

$$\begin{aligned} \Delta N^L = & p_0^- p_{b=\uparrow}^- P_{0 \rightarrow \uparrow} + p_0^- p_{b=\downarrow}^- P_{0 \rightarrow \downarrow} \\ & - p_{\uparrow}^- p_{b=\uparrow}^- (P_{\uparrow \rightarrow 0} + P_{\uparrow \rightarrow \downarrow}) + p_{\uparrow}^- p_{b=\downarrow}^- P_{\downarrow \rightarrow \uparrow} \\ & - p_{\downarrow}^- p_{b=\uparrow}^- (P_{\downarrow \rightarrow 0} + P_{\downarrow \rightarrow \downarrow}) + p_{\downarrow}^- p_{b=\downarrow}^- P_{\downarrow \rightarrow \uparrow}. \end{aligned} \quad (9)$$

Here, $P_{s \rightarrow s'}$ denotes the transition probability from the system state s to s' due to the Liouvillian \mathcal{W} . For $\tau \rightarrow \infty$ this becomes $P_{s \rightarrow s'} = \bar{p}_{s'}$ with $\bar{p}_{s'}$ from Eq. (3). Parametrizing the bit probabilities by the excess parameter δ^- yields after some manipulation for ΔN^L and, following a similar procedure, for ΔN^R :

$$\Delta N^L = p_0^- P_{0 \rightarrow \uparrow} \quad (10)$$

$$+ p_1^- \left[\frac{1 - \delta^-}{2} P_{\downarrow \rightarrow \uparrow} - \frac{1 + \delta^-}{2} (P_{\uparrow \rightarrow 0} + P_{\uparrow \rightarrow \downarrow}) \right],$$

$$\Delta N^R = p_0^- P_{0 \rightarrow \downarrow} \quad (11)$$

$$+ p_1^- \left[\frac{1 + \delta^-}{2} P_{\uparrow \rightarrow \downarrow} - \frac{1 - \delta^-}{2} (P_{\downarrow \rightarrow 0} + P_{\downarrow \rightarrow \uparrow}) \right].$$

It is possible to verify particle number conservation $\Delta N^L + \Delta N^R = 0$ and for later purposes we note the following identity:

$$2\Delta N^L = \Delta N^L - \Delta N^R = (p_0^- - 1)\delta^- + p_{\uparrow}^- - p_{\downarrow}^-. \quad (12)$$

The average current is defined by $I^L(\tau) \equiv \frac{\Delta N^L(\tau)}{\tau}$. Two particularly simple expressions are obtained in the limits

$$\lim_{V \rightarrow 0} \lim_{\tau \rightarrow \infty} \Delta N^L = -\frac{\delta^-}{3}, \quad \lim_{V \rightarrow 0} \lim_{\tau \rightarrow 0} I^L = -\frac{\delta^-}{6}. \quad (13)$$

The minus sign indicates that an excess of spin up electrons on the tape ($\delta^- > 0$) generates a current from right to left.

Finally, we compute the change in the statistics of the outgoing bits. The excess parameter of the tape after the interaction with the system is given by

$$\delta^+ = p_{b=\uparrow}^+ - p_{b=\downarrow}^+ = p_0^- \delta^- + p_{\uparrow}^- - p_{\downarrow}^-. \quad (14)$$

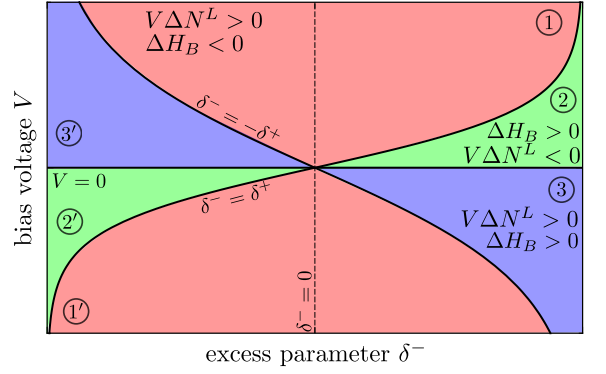


FIG. 3: Phase diagram of the system over δ^- and V showing the different working modes: information eraser mode (regions 1 and 1', red color), Maxwell demon mode (regions 2 and 2', green color) and the 'third mode' (regions 3 and 3', blue color). The primed regions are connected to the unprimed regions by the symmetry relation. For clarity we also plot the line $\delta^- = 0$ (dashed). Note that the line $\delta^+ = -\delta^-$ changes with τ but the overall structure remains the same.

Using Eq. (12) we can write

$$\delta^+ = \delta^- + 2\Delta N^L, \quad (15)$$

an identity we will use later on prove a second law like inequality [47]. Having δ^+ allows us to compute the change in the Shannon entropy of the bits defined by [48]

$$\Delta H_B \equiv H[\delta^+] - H[\delta^-] \quad (16)$$

with

$$H[\delta] \equiv -\frac{1+\delta}{2} \ln \frac{1+\delta}{2} - \frac{1-\delta}{2} \ln \frac{1-\delta}{2}. \quad (17)$$

If $\Delta H_B > 0$ we effectively *write* information to the tape, i.e., increase its entropy, while if $\Delta H_B < 0$ implies that we *erase* information from the tape effectively lowering its entropy.

B. Working Modes of the System

In principle, one can distinguish three different working modes in which our device can operate. First, it can be used to pump electrons against the bias (i.e., $V\Delta N^L < 0$) while simultaneously writing information on the tape ($\Delta H_B > 0$) what we will call the 'Maxwell demon mode'. Second, one can use it to erase information ($\Delta H_M < 0$) by exploiting the flow of electrons with the bias ($V\Delta N^L > 0$) ('eraser mode'). Third, we can also have $V\Delta N^L > 0$ and $\Delta H_B > 0$ where electrons flow with the bias while simultaneously information is written on the tape [49]. A possible working mode where $V\Delta N^L < 0$ and $\Delta H_B < 0$ is forbidden by the second law of thermodynamics as we will see in the next section.

These modes are summarized in Fig. 3 and are separated from each other by three lines in the phase diagram. One line is characterized by $V = 0$ and the other

two lines are given by $\Delta H_B = 0$, which can happen either if $\delta^+ = \delta^-$ or if $\delta^+ = -\delta^-$. The first case is given by

$$\delta^+ = \delta^- \Leftrightarrow \delta^- = \frac{e^V - 1}{e^V + 1} = \tanh \frac{V}{2} \quad (18)$$

where $\tanh(x)$ denotes the hyperbolic tangent. This line is independent of τ . By contrast, the line for which $\delta^+ = -\delta^-$ depends on τ and has in general a more complicated functional relation.

The total phase space $V \times \delta^-$ for a given τ is divided into six regions, of which only three are independent due to the symmetry. Specifically, we have the relations

$$\Delta N^L(V, \delta^-) = \Delta N^R(-V, -\delta^-), \quad (19)$$

$$\delta^+(V, \delta^-) = -\delta^+(-V, -\delta^-), \quad (20)$$

$$\Delta H_B(V, \delta^-) = \Delta H_B(-V, -\delta^-). \quad (21)$$

C. Second Law

In appendix B we obtain the following second law like inequality:

$$\Delta_i S_{\text{tape}} \equiv V \Delta N^L + \Delta H_B = -\Delta W + \Delta H_B \geq 0 \quad (22)$$

using the equivalence between the spin valve and the Brownian ratchet model ($V \Delta N^L = -\Delta W$). We will call $\Delta_i S_{\text{tape}}$ the entropy production (per cycle) of the tape model. In fact, from the point of view of the system there are different versions of the second law possible as we will see in the next section. Thus, it is hard to justify speaking about *the* second law for our device.

IV. MEASUREMENT BASED FEEDBACK SCHEME

We now turn to a closed-loop or feedback control scheme, which relies on explicit measurements and subsequent control actions depending on the measurement outcomes. This approach is closer to the original idea of Maxwell. In fact, our inclusive approach above has nothing to do with *feedback* control because we do not use the information we are writing on the tape to change the system dynamics. Moreover, the ‘information’ on the tape does not directly represent information about the state of the system, but only about ΔN_L , a quantity which is rather related to the reservoir than to the system.

A. Measurement and Feedback

We assume we have a 1-bit memory M with state space $\mathcal{M} = \{N, S\}$ where N, S will specify the action of the feedback later on [50]. Prior to the measurement the combined probability distribution is given by

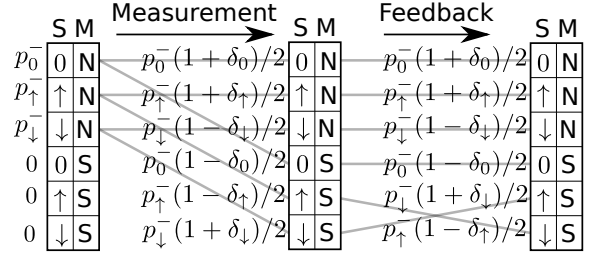


FIG. 4: Probabilities to find the system and the memory in their respective states prior to the measurement, after the measurement and after the feedback. The possible transitions are marked with shaded grey lines.

$p(s, m) = p(s)p(m) = p_s^- \delta_{mN}$ where we assume that the state of the memory is initially in $m = N$ and we denote $p(s) = p_s^-$ as in the previous section. We then perform a measurement of the quantum dot spin valve: ideally, if the state of the dot is $s \in \{0, \uparrow\}$ we leave the memory in the state N , but if the state of the dot is $s = \downarrow$ we set the memory to the state S . In general there will be a measurement error, which can be characterized by the conditional probability $P[m|s]$ to find the memory in the state m given the system was in state s . The state after the measurement is then given by $p'(s, m) = P[m|s]p_s^-$ with $\sum_m P[m|s] = 1$. We parametrize the conditional probabilities as [51]

$$P[S|0] = \frac{1-\delta_0}{2}, \quad P[S|\uparrow] = \frac{1-\delta_\uparrow}{2}, \quad P[N|\downarrow] = \frac{1-\delta_\downarrow}{2}. \quad (23)$$

Finally, after the measurement we flip the spin of the quantum dot, if the memory is in state S (‘switch’) or we do nothing if the memory is in state N (‘no switch’). Ideally, because the energy for spin up and down in the quantum dot are assumed to be equal, this feedback operation requires no energy expenditure. Of course, if the state of the system is 0, then nothing happens. The combined probability distribution of system and memory after the feedback can be inferred from Fig. 4. The state of the quantum dot spin valve is now given by

$$\rho_S^+ = \begin{pmatrix} p_0^- & \\ p_\uparrow^- \frac{1+\delta_\uparrow}{2} + p_\downarrow^- \frac{1+\delta_\downarrow}{2} & \\ p_\downarrow^- \frac{1-\delta_\downarrow}{2} + p_\uparrow^- \frac{1-\delta_\uparrow}{2} & \end{pmatrix}. \quad (24)$$

To obtain identical system dynamics compared to the situation involving the tape of bits, we only have to assure that Eq. (8) coincides with Eq. (24) because the time evolution in between two successive measurement and feedback steps is the same as before, see Eq. (7). Thus, we will from now on choose

$$\delta_\uparrow = \delta_\downarrow \equiv \delta^-. \quad (25)$$

All quantities associated to the system (such as ΔN^L) are the same for both approaches. What makes the two approaches different is the form of information processing and the related second laws.

B. Second Law

The second law of thermodynamics for feedback controlled systems predicts [12, 37, 38]

$$\Delta_i S_{fb} \equiv V \Delta N^L + I(S; M') \geq 0, \quad (26)$$

which is equivalent to $\Delta W \leq I(S; M')$ due to our mapping to the Brownian ratchet model. Here, $I(S; M')$ is the mutual information, which measures the amount of correlation between the system state before the feedback and the post-measurement state of the memory. It can be defined as [39]

$$I(S; M') \equiv H_{M'} - H_{M'|S} \quad (27)$$

where $H_{Y|X} \equiv -\sum_x p(x) \sum_y p(y|x) \ln p(y|x)$ is the conditional entropy of Y given X . The mutual information is always non-negative, $I(X; Y) \geq 0$, and vanishes only if $p(x, y) = p(x)p(y)$ [39].

More explicitly, the mutual information can be written in our case as

$$I(S; M') = -\sum_m p'(m) \ln p'(m) + \sum_s p_s^- \sum_m P[m|s] \ln P[m|s] \quad (28)$$

where $p'(m) = \sum_s p'(m, s) = \sum_s P[m|s] p_s^-$ is the marginal probability distribution of the memory after the measurement and the conditional probabilities $P[m|s]$ were introduced in Eq. (23), see also Fig. 4. Together with Eq. (25) and parametrizing the marginal probability distribution as $p'(m = N) = (1 + \delta_M)/2$ and $p'(m = S) = (1 - \delta_M)/2$, we can write the mutual information as

$$I(S; M') = H[\delta_M] - p_0^- H[\delta_0] - p_1^- H[\delta^-] \quad (29)$$

with $H[x]$ defined as in Eq. (17). Here, δ_M is given by

$$\delta_M = \delta_0 p_0^- + \delta^- (p_\uparrow^- - p_\downarrow^-). \quad (30)$$

This equation is analogous to Eq. (14).

C. Comparison of the two Second Laws

We wish to compare the second law from our inclusive approach, Eq. (22), with the second law for feedback controlled system, Eq. (26), using the mutual information from Eq. (29). First of all, we note that one can not address problems like information erasure within the feedback approach because the mutual information is always positive. Hence, we will focus on the Maxwell demon regime, i.e., region 2 in Fig. 3. Furthermore, we see that the mutual information still depends on δ_0 although the dynamics of the system are completely *independent* of δ_0 . For a better comparison we thus find it sensible

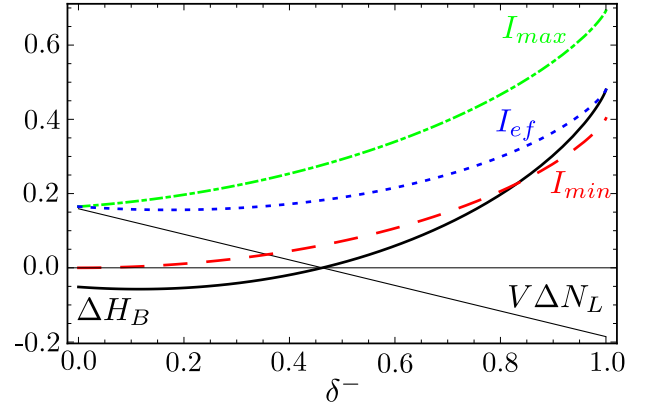


FIG. 5: Plot of the delivered work $-\Delta W = V \Delta N^L$ (black thin line), the change in the Shannon entropy of the bits ΔH_B (black thick line) and the maximum I_{\max} (green dash-dotted line), the minimum I_{\min} (red dashed line) and the error-free mutual information I_{ef} (blue dotted line) over the excess parameter δ^- . We see that the sum of $V \Delta N^L$ and any of the information theoretic quantities ΔH_B , I_{\max} , I_{\min} or I_{ef} is always positive as imposed by Eqs. (22) and (26). We further chose $V = 1$ and $\tau = 10$.

to compare the case $\delta_0 = 1$, which corresponds to an error-free measurement of the empty dot state, with the maximum and minimum value of the mutual information in dependence of the parameter δ_0 . These are attained for

$$\delta_0^{\max} = -1, \quad \delta_0^{\min} = \delta^- \frac{p_\uparrow^- - p_\downarrow^-}{p_\uparrow^- + p_\downarrow^-} \quad (31)$$

as we establish in appendix C. The mutual information (29) in these three cases becomes after using Eq. (30)

$$I_{\max} \equiv I(S; M')(\delta_0^{\max}) \quad (32)$$

$$= H[-p_0^- + \delta^- (p_\uparrow^- - p_\downarrow^-)] - p_1^- H[\delta^-],$$

$$I_{\text{ef}} \equiv I(S; M')(\delta_0 = 1) \quad (33)$$

$$= H[p_0^- + \delta^- (p_\uparrow^- - p_\downarrow^-)] - p_1^- H[\delta^-],$$

$$I_{\min} \equiv I(S; M')(\delta_0^{\min}) \quad (34)$$

$$= H[\delta_0^{\min}] - p_0^- H[\delta_0^{\min}] - p_1^- H[\delta^-]$$

where the subscript ‘ef’ stands for ‘error-free’. The maximum amount of correlation between system and memory is thus reached for $\delta_0 = -1$, which indeed would also correspond to an error-free measurement because a measurement, which always yields the wrong outcome, is very reliable, too. On the other hand, the dynamically irrelevant parameter δ_0 gives us the possibility to decrease the amount of correlation such that we obtain a stronger bound in Eq. (26) on the maximum extractable work ΔW .

Numerical results are presented in Fig. 5 showing that neither the second law for the inclusive approach nor the second law for feedback control always provides a tighter bound on the maximum amount of extractable work.

Thus, there is probably no trivial relationship connecting the two different second laws for arbitrary δ^- . Even for the special case of a fully ordered tape, i.e., for $\delta^- = 1$, there seems to be no general relation, but we can easily confirm $I_{\min} \leq \Delta H_B = I_{\text{ef}} \leq I_{\max}$.

V. EFFECTIVE MASTER EQUATION

A. Poisson distributed Tape

Previously the interaction with the bit occurred at regular intervals τ , i.e., the so called waiting time distribution was given by $p_t = \delta(t - \tau)$. Instead, we now assume an exponential waiting time distribution $p_t = \gamma e^{-\gamma t}$, which implies that the number of events (interactions with a bit) in a fixed interval is Poisson distributed. Here, the parameter $\gamma > 0$ describes the strength of the Poisson process. Although we speak of ‘interaction with a bit’, the resulting ME for the system is the *same* as if we would use the feedback scheme from the previous section with Poisson distributed measurement times.

Due to the Poisson distributed tape we can now describe the evolution of the system by a single differential equation as opposed to the discrete maps (7) and (8). The resulting effective ME is derived in appendix D and can be written as $\partial_t \rho_S(t) = \mathcal{W}_{\text{eff}} \rho_S(t)$ with $\mathcal{W}_{\text{eff}} = \mathcal{W}_L + \mathcal{W}_R + \mathcal{W}_B$ where \mathcal{W}_L and \mathcal{W}_R are given in Eq. (1) and

$$\mathcal{W}_B = \frac{\gamma}{2} \begin{pmatrix} 0 & 0 & 0 \\ 0 & -(1 - \delta^-) & 1 + \delta^- \\ 0 & 1 - \delta^- & -(1 + \delta^-) \end{pmatrix}. \quad (35)$$

The contribution \mathcal{W}_B due to the interaction with the information reservoir, i.e., the tape of bits, has an apparent non-thermal character and is responsible for modifications of thermodynamic relations even in the absence of information processing or feedback [40]. The effective rate matrix \mathcal{W}_B could, however, also result from coarse-graining an underlying microscopic model, which has a standard thermodynamic interpretation (as in Ref. [9]).

We see that the effective Liouvillian possesses a clear separation into contributions from three different ‘reservoirs’, and thus can be examined using standard thermodynamics. In this picture, each reservoir induces certain transitions from σ to σ' with corresponding currents $I[\sigma \rightarrow \sigma']$. We abbreviate $I^L = I[0 \rightarrow \uparrow]$, $I^R = I[0 \rightarrow \downarrow]$ and $I^B = I[\uparrow \rightarrow \downarrow] = \frac{\gamma}{2}[(1 - \delta^-)p_{\uparrow} - (1 + \delta^-)p_{\downarrow}]$, which are related at steady state via

$$I^L + I^R = 0, \quad I^L - I^B = 0, \quad I^R + I^B = 0. \quad (36)$$

Thus, it suffices to know only one current, e.g. I^L , to deduce the value of all other currents (‘tight coupling’). Note that the heat flow between the bits and system is zero because the states \uparrow and \downarrow have the same energy. Furthermore, I^B describes the net rate of spin flips (from

\uparrow to \downarrow) in the system, which equals the net rate of bit flips (now from \downarrow to \uparrow) in the tape.

The entropy production associated with the rate matrix \mathcal{W} taking care of the splitting into the different reservoir contributions becomes

$$\dot{S}_i = (V + \mathcal{A})I^L \geq 0, \quad \mathcal{A} \equiv \ln \frac{1 - \delta^-}{1 + \delta^-}. \quad (37)$$

Following Refs. [9, 21, 22] we now interpret $\mathcal{A}I^L$ as an effective information current because it shows up as a new term in the entropy production due to the interaction with the tape, and \mathcal{A} is an information or feedback affinity depending only on the parameters describing the state of the incoming bits (or the measurement error in the feedback scheme). A priori however it is not quite clear how the effective information current is related to information theoretic quantities as, e.g., entropy or mutual information. While progress has been made toward clarifying this issue for bipartite systems [27, 28, 41, 42], our model does not have a bipartite structure due to the simultaneous change of the system and bit state. Nevertheless, within our model, we can indeed show that (see appendix E)

$$\frac{dH_B(t)}{dt} = \mathcal{A}I^L, \quad (38)$$

that is to say the effective information current introduced *ad hoc* coincides with the change of Shannon entropy of an idealized tape of bits, which induces the transitions between \uparrow and \downarrow . Thus, Eq. (37) becomes

$$\dot{S}_i = VI^L + \frac{dH_B(t)}{dt} \geq 0 \quad (39)$$

In this sense, the two second laws (22) and (39) present the same physical statement.

Note that the situation in which the effective ME (35) is valid is not as easy to compare with the previous situation of constant time intervals, as was the case for the two level system studied in Ref. [26]. In Ref. [26] the state of the system after the arrival of the new bit or after the feedback was *independent* of the system state before. This allowed them to average quantities obtained for the constant interval case by the exponential distribution $\gamma e^{-\gamma t}$ to obtain the corresponding quantities for the Poisson distributed case. In contrast, in our case the state of the system after the swap operation or after the feedback is still dependent on the state of the system before, see Eqs. (8) and (24). Thus, averaging previous quantities obtained in Secs. III and IV with $\gamma e^{-\gamma t}$ is not meaningful, as they do not coincide with any of the quantities obtained from Eq. (35).

However, we expect the case of constant intervals (Secs. III and IV) to be comparable with the case of Poisson distributed intervals, if we let the size of the intervals go to zero, i.e., $\tau \rightarrow 0$ or $\gamma \rightarrow \infty$. We will call this limit the ‘infinite fast feedback limit’.

B. Infinite fast Feedback Limit

In the infinite fast feedback limit it is again possible to derive a new effective Liouvillian $\tilde{\mathcal{W}}_{\text{eff}} = \tilde{\mathcal{W}}_L + \tilde{\mathcal{W}}_R$ by coarse graining the fast dynamics associated with \mathcal{W}_B . This new Liouvillian governs the time evolution of the reduced probability vector (p_0, p_1) with $p_1 = p_\uparrow + p_\downarrow$ and is given by

$$\begin{aligned}\tilde{\mathcal{W}}_L &= \Gamma \begin{pmatrix} -f_L & (1-f_L)(1+\delta^-) \\ f_L & -(1-f_L)(1+\delta^-) \end{pmatrix}, \\ \tilde{\mathcal{W}}_R &= \Gamma \begin{pmatrix} -f_R & (1-f_R)(1-\delta^-) \\ f_R & -(1-f_R)(1-\delta^-) \end{pmatrix}.\end{aligned}\quad (40)$$

We have thus effectively traced out the interaction with the ‘bit reservoir’ but obtained modified rates for the electrons to tunnel out of the dot. This modification of the rates, which yields a modified local detailed balance relation, was interpreted as a signature of an idealized Maxwell demon in [9, 20, 21]. We remark that it is also possible to deduce this ME directly, without invoking a Poisson distribution of the bits but by starting with Eqs. (7) and (8) and expanding the generator to first order in τ : $e^{\mathcal{W}\tau} = 1 + \mathcal{W}\tau + \mathcal{O}(\tau^2)$.

The thermodynamic analysis of Eq. (40) again follows the standard procedure. We now have only two currents, which must balance: $\tilde{I}^L + \tilde{I}^R = 0$. Explicitly we have

$$\tilde{I}^L = -\frac{\delta^- + (\delta^- - 1)e^V + 1}{\delta^- - (\delta^- - 3)e^V + 6e^{V/2} + 3}. \quad (41)$$

The second law takes on the same form as Eq. (39), but is numerically different because it contains the current \tilde{I}^L and not I^L :

$$\dot{\tilde{S}}_i = (V + \mathcal{A})\tilde{I}^L \geq 0. \quad (42)$$

As a first crosscheck of the validity of Eq. (40), any computer algebra program can confirm

$$\lim_{\tau \rightarrow 0} \frac{\Delta N^L}{\tau} = \lim_{\gamma \rightarrow \infty} I^L = \tilde{I}^L \quad (43)$$

where ΔN^L was defined in Eq. (10). Together with Eq. (38) it should be also immediately clear that

$$\lim_{\tau \rightarrow 0} \frac{(\Delta_i S_{\text{tape}})}{\tau} = \lim_{\gamma \rightarrow \infty} \dot{S}_i = \dot{\tilde{S}}_i. \quad (44)$$

By contrast, the mutual information (29) becomes constant (and non-zero) for $\tau \rightarrow 0$ such that its rate diverges: $\lim_{\tau \rightarrow 0} I(S; M')/\tau = \infty$. Thus, the second law involving discrete feedback, Eq. (26), no longer yields a useful bound. This does not contradict the claim that the mutual information can give a tighter bound for the error-free case, see Sec. IV C, because for $\delta^- \rightarrow 1$ the feedback affinity \mathcal{A} also becomes unbounded. Results are plotted in Fig. 6.

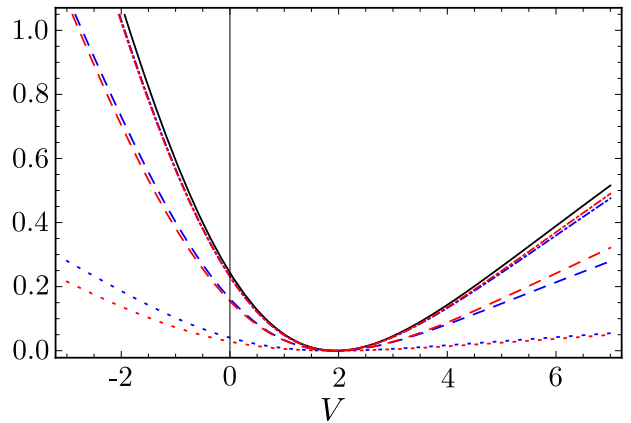


FIG. 6: Plot of the entropy production rates \dot{S}_i [Eq. (42), black solid line] as well as \dot{S}_i [Eq. (39), blue lines] and $\Delta_i S_{\text{tape}}/\tau$ [Eq. (22), red lines] for $\gamma = 0.1, \tau = 10$ (dotted lines), $\gamma = 1, \tau = 1$ (dashed lines) and $\gamma = 10, \tau = 0.1$ (dash-dotted lines) over the bias voltage V . We chose $\delta^- = 0.75$.

VI. DISCUSSION

We now briefly summarize and discuss our results. We divide the discussion into three main parts, which can be related to Secs. II and III (i), Sec. IV (ii) and Sec. V (iii).

(i) We have constructed a solvable physical model able to rectify thermal fluctuations to transport electrons against a bias (i.e., to charge a battery) by writing information on a tape without energy expenditure. This ‘information-electric device’ is thermodynamically equivalent to the model in Ref. [10]. The physical treatment of the problem allows us to clearly state the assumptions needed to obtain this ideal equivalence. First, to have an unchanged first law of thermodynamics, we need to assume a frictionless gliding tape of bits and degenerate spin eigenstates in the spin valve and on the tape. Second, to keep the discussion at a moderate level, we have neglected various experimentally important imperfections, which would reduce the overall efficiency of our device. Namely, we have assumed perfectly opposite polarized leads, neglected phonon interactions and assumed a fully unitary and error-free swap gate. Although we used the physical picture of a spin valve interacting via a Heisenberg Hamiltonian with the bits, it is worth pointing out that this could be also realized with other models: aside from the Brownian ratchet analogue, one could, for instance, use double quantum dots where the left and right eigenstates form a pseudospin giving rise to the same system dynamics, and – as pointed out at the end of Sec. II C – one does not necessarily need to implement a tape of bits although this picture might still be helpful for thermodynamic considerations. Finally, it is interesting to note the similarity of these information driven devices with the theory of the micromaser [43] in which a stream of excited two-level atoms (like a tape of bits) is injected into a high quality cavity building up a photon field. In

deed, the device proposed by Scully in Ref. [44] can be regarded as a quantum optical version of an information driven engine, where the entropy of the motional degrees of freedom of the two-level atoms is increased while energy is continuously extracted from the population of the excited states.

(ii) We have formulated a discrete measurement and feedback model with identical system dynamics but a conceptually different second law of thermodynamics. Indeed, there are multiple second laws due to the fact that a dynamically irrelevant parameter (δ_0) entered the mutual information. This additional parameter might be more rigorously exploited to give a stronger bound on the extractable work by minimizing the mutual information with respect to δ_0 , although its precise value δ_0^{\min} might not be physically intuitive [Eq. (31)]. Furthermore, although we used identical ‘information resources’ (namely a two-state memory equivalent to a bit on the tape), the second law for discrete feedback can provide a smaller, equal or larger bound than the second law involving the tape of bits, hence it remains a challenge to treat the different forms of information processing in one unified framework.

(iii) We have derived an effective ME governing the system evolution for Poisson distributed events where an ‘event’ can refer either to an interaction with a bit or a measurement and feedback step. The ME is the same for both approaches and we believe that its effective entropy production coincides with the entropy production of the tape model. *A priori* this is not obvious, because we have seen that the mutual information can coincide with the change in Shannon entropy of the bits [see Sec. IV

or point (ii)]. However, we have shown that the *a posteriori* postulated information current equals the rate of entropy change of an idealized tape of bits, see Eq. (38), thus giving the phenomenologically introduced information current a precise physical and information theoretic basis. Furthermore, in the limit of infinite fast feedback we have demonstrated that the effective entropy production coincides with the entropy production of the tape model whereas the rate of mutual information diverges. It is worth mentioning two more things: first of all, the information current (38) has the standard form often encountered in thermodynamics of a current times an affinity. Thus, the finding in Ref. [26] – that the entropy production involving the Shannon entropy difference of the information reservoir has an *extra* term different from the usual current times affinity – is in general not true. Second, we want to mention that under certain assumptions it is indeed possible to define a rate of mutual information as was done, for instance, in Ref. [12]. Note, however, that this has not the standard form of a current times an affinity and is always positive and thus, makes it again impossible to address question of information erasure or related issues.

Acknowledgments

PS thanks Sebastian Deffner and Zhiyue Lu for many stimulating discussions. Financial support by the DFG (SCHA 1646/2-1, SFB 910, and GRK 1558) is gratefully acknowledged. CJ gratefully acknowledges financial support from the National Science Foundation (USA) under grant DMR-1206971.

-
- [1] J. C. Maxwell, *Theory of Heat* (1871).
 - [2] R. Landauer, IBM J. Res. Dev. **5**, 183 (1961).
 - [3] C. H. Bennett, Int. J. Theor. Phys. **21**, 905-940 (1982).
 - [4] L. Szilard, Zeitschrift für Physik **53**, 840-856 (1929).
 - [5] M. Smulochowski, Zeitschrift für Physik **13**, 1069-1080 (1912).
 - [6] R. P. Feynman, R. Leighton, and M. Sands, *The Feynman Lectures on Physics* (Addison-Wesley, Reading, 1963).
 - [7] J. D. Norton, Entropy **15**, 4432-4483 (2013).
 - [8] P. A. Skordos and W. H. Zurek, Am. J. Phys. **60**, 876 (1992).
 - [9] P. Strasberg, G. Schaller, T. Brandes, and M. Esposito, Phys. Rev. Lett. **110**, 040601 (2013).
 - [10] D. Mandal and C. Jarzynski, Proc. Natl. Acad. Sci. **109**, 11641 (2012).
 - [11] A. C. Barato and U. Seifert, Europhys. Lett. **101**, 60001 (2013).
 - [12] J. M. Horowitz, T. Sagawa, and J. M. R. Parrondo, Phys. Rev. Lett. **111**, 010602 (2013).
 - [13] D. Mandal, H. T. Quan, and C. Jarzynski, Phys. Rev. Lett. **111**, 030602 (2013).
 - [14] S. Deffner and C. Jarzynski, Phys. Rev. X **3**, 041003 (2013).
 - [15] S. Deffner, Phys. Rev. E **88**, 062128 (2013).
 - [16] J. Hoppenau and A. Engel, Euro. Phys. Lett. **105**, 50002 (2014).
 - [17] S. Datta, *Nanodevices and Maxwell's Demon*, Lecture Notes, arXiv: 0704.1623.
 - [18] Z. Lu, D. Mandal, and C. Jarzynski, Physics Today **67**(8), 60 (2014).
 - [19] H. M. Wiseman and G. J. Milburn, *Quantum Measurement and Control* (Cambridge University Press, Cambridge, 2010).
 - [20] G. Schaller, C. Emary, G. Kiesslich, and T. Brandes, Phys. Rev. B **84**, 085418 (2011).
 - [21] M. Esposito and G. Schaller, Euro. Phys. Lett. **99**, 30003 (2012).
 - [22] P. Strasberg, G. Schaller, T. Brandes, and M. Esposito, Phys. Rev. E **88**, 062107 (2013).
 - [23] T. Munakata and M. L. Rosinberg, J. Stat. Mech., P05010 (2012).
 - [24] T. Munakata and M. L. Rosinberg, J. Stat. Mech., P06014 (2013).
 - [25] H. Tasaki, arXiv: 1308.3776.
 - [26] A. C. Barato and U. Seifert, Phys. Rev. Lett. **112**, 090601 (2014).
 - [27] J. M. Horowitz and M. Esposito, Phys. Rev. X **4**, 031015 (2014).

- (2014).
- [28] N. Shiraishi and T. Sagawa, arXiv: 1403.4018.
 - [29] M. Bauer, A. C. Barato, and U. Seifert, J. Stat. Mech., P09010 (2014).
 - [30] L. Bogani and W. Wernsdorfer, Nat. Mater. **7**, 179 (2008).
 - [31] M. Urdampilleta, S. Klyatskaya, J.-P. Cleuziou, M. Ruben, and W. Wernsdorfer, Nat. Mater. **10**, 502 (2011).
 - [32] M. Braun, J. König, and J. Martinek, Phys. Rev. B **70**, 195345 (2004).
 - [33] P. Fulde, *Electron Correlations in Molecules and Solids* (Springer-Verlag, 3rd. Edition, Berlin, 1995).
 - [34] D. Loss and D. P. DiVincenzo, Phys. Rev. B **57**, 120 (1998).
 - [35] J. R. Petta *et al.*, Science **309**, 2180 (2005).
 - [36] S. Toyabe, T. Sagawa, M. Ueda, E. Muneyuki, and M. Sano, Nat. Phys. **6**, 988 (2010).
 - [37] F. J. Cao and M. Feito, Phys. Rev. E **79**, 041118 (2009).
 - [38] S. Deffner and E. Lutz, arXiv: 1201.3888.
 - [39] T. M. Cover and A. J. Thomas, *Elements of Information Theory* (Wiley-Interscience, Hoboken, NJ, 2006).
 - [40] K. Kanazawa, T. Sagawa, and H. Hayakawa, Phys. Rev. E **87**, 052124 (2013).
 - [41] G. Diana and M. Esposito, J. Stat. Mech., P04010 (2014).
 - [42] D. Hartich, A. C. Barato, and U. Seifert, J. Stat. Mech., P02016 (2014).
 - [43] M. O. Scully and M. Suhail Zubairy, *Quantum Optics* (Cambridge University Press, Cambridge, 1997).
 - [44] M. O. Scully, Phys. Rev. Lett. **87**, 220601 (2001).
 - [45] R. Kosloff, Entropy **15**, 2100-2128 (2013).
 - [46] To obtain Eq. (1) from Ref. [32] one needs to take the limits (in their notation) of (i) ultra strong Coulomb blockade ($U \rightarrow \infty$), (ii) oppositely polarized leads ($\hat{\mathbf{n}}_L = \mathbf{e}_x = -\hat{\mathbf{n}}_R$), (iii) fully polarized leads ($p_L = p_R = 1$) and (iv) for simplicity equal tunneling rates $\Gamma_L = \Gamma_R \equiv \Gamma/2$.
 - [47] Also compare this expression with the equivalent Eqs. [S17] and [3] in Ref. [10]. The equivalence is obvious by noting that the circulation Φ in Ref. [10] corresponds to the change in particle number ΔN^L in our case, where Φ is proportional to the extracted work W in Ref. [10] in the same way as ΔN^L is proportional to the extracted work in our Brownian ratchet version.
 - [48] Note that this definition ignores correlations in the outgoing bit stream, which are present but rather small in our case. For $\tau \rightarrow \infty$ these correlations vanish, because the system has time to relax to the steady state (3), which is uncorrelated from the state of the bit. Also see, for instance, the discussion in Ref. [10].
 - [49] Note that we always have $V\Delta N^L = 0$ if the spin valve does not interact with a tape. Thus, in contrast to Ref. [10] it is probably better not to call this third mode a 'dud' since it can still accomplish a useful task.
 - [50] It is easy to think of memories with more than two states, but in a certain sense a two state memory is minimal and could be also realized by a bit in our tape model. Thus, we will focus on the two state memory.
 - [51] Very often one uses a parameter $\epsilon_s \in [0, 1]$ ($s \in \{0, \uparrow, \downarrow\}$) to quantify the measurement error, but for better comparison with the discussion in Sec. III we use the excess parameter $\delta_s \equiv 1 - 2\epsilon_s$ instead, which now quantifies the excess of correct measurements compared to faulty ones (for $\delta_s > 0$).

Appendix A: Proof of $\Delta W_{\text{pull}} = 0$

We denote the combined system and bit Hamiltonian during one swap operation by $H_{SB}(t) \equiv H_S + H_B + V(t)$ where $t \in [t_-, t_+]$ with $t_+ = t_- + \delta t$ and V denotes the system-bit coupling generated by the Heisenberg interaction fulfilling $V(t_-) = 0 = V(t_+)$. The change in system and bit energy is given by

$$\begin{aligned} \frac{d}{dt}E(t) &= \frac{d}{dt}\text{tr}_{SB}[H_{SB}(t)\rho_{SB}(t)] \\ &= \text{tr}_{SB}[\dot{H}_{SB}(t)\rho_{SB}(t) + H_{SB}(t)\dot{\rho}_{SB}(t)] \\ &\equiv \dot{W}_{\text{pull}}(t) + \dot{Q}(t). \end{aligned} \quad (\text{A1})$$

In the general theory of open quantum systems the first term on the right hand side is interpreted as the work and the second as the heat flow [45]. Because we assume the system to change unitarily we have

$$\begin{aligned} \dot{Q} &= \text{tr}_{S+B}[H_{S+B}(t)\dot{\rho}_{S+B}(t)] \\ &= -i\text{tr}\{H_{S+B}(t)[H_{S+B}(t), \rho(t)]\} = 0. \end{aligned} \quad (\text{A2})$$

Thus, after integrating over one swap operation from t_- to t_+ , we have $E(t_+) - E(t_-) = \Delta W_{\text{pull}}$ and hence,

$$\begin{aligned} \Delta W_{\text{pull}} &= \text{tr}_{SB}[H_{SB}(t_+)\rho_{SB}(t_+) - H_{SB}(t_-)\rho_{SB}(t_-)] \\ &= \text{tr}_{SB}[\{H_S(t_+) + H_B(t_+)\}\rho_{SB}(t_+) \\ &\quad - \text{tr}_{SB}[\{H_S(t_-) + H_B(t_-)\}\rho_{SB}(t_-)] \\ &= 0 \end{aligned} \quad (\text{A3})$$

because $V(t_-) = 0 = V(t_+)$. Note that these results remains true even if we introduce a certain asymmetry Δ in the energy of spin up and down for both system and bit Hamiltonian [i.e., $H_{S/B} = \epsilon_{s/b}|\uparrow\rangle\langle\uparrow| + (\epsilon_{s/b} + \Delta)|\downarrow\rangle\langle\downarrow|$], but this would eventually affect the first law of thermodynamics for the system (or bit) alone.

Appendix B: Proof of $\Delta_i S_{\text{tape}} \geq 0$

Deducing the second law for our model is not as clear as one might expect, especially because our dynamics are mixed between a unitary evolution and a subsequent dissipative (ME like) evolution. Arguments making use of relative entropy (or the Kullback Leibler divergence) as in Ref. [13] or arguments relying on an integral fluctuation theorem do not apply as easily to our setup. We will thus explicitly prove the second law in the same way as done by Mandal and Jarzynski [10]. The proof proceeds in two steps: first, we will show Eq. (22) for the stationary case $\tau \rightarrow \infty$ and we will then use this result to prove it for finite τ .

Note that for our proof we only have to focus on the regions 1 and 2 in Fig. 3. In region 3 we have $V\Delta N^L > 0$ and $\Delta H_B > 0$ such that Eq. (22) is trivially fulfilled and regions 1', 2' and 3' follow from regions 1,2 and 3 by symmetry.

The advantage of the long time limit $\tau \rightarrow \infty$ is that many expressions become quite simple. In particular, we have

$$\Delta N^L(\infty) = -\frac{V}{2} \frac{\delta^- + 1 + e^V(\delta^- - 1)}{1 + e^{V/2} + e^V}, \quad (\text{B1})$$

$$\delta^+(\infty) = \frac{e^V - 1 + e^{V/2}\delta^-}{1 + e^{V/2} + e^V}. \quad (\text{B2})$$

where $f(\infty) \equiv \lim_{\tau \rightarrow \infty} f(\tau)$ for any function f of τ (if the limit exists). To simplify the subsequent algebra it is convenient to rescale the variables. First, we introduce $x \equiv e^{V/2}$ such that $x \in [1, \infty)$ for V in region 1,2. Next, we parametrize δ^- by another parameter h through

$$\delta^- = \tanh \frac{V}{2} + \frac{h}{e^V + 1} = \frac{x^2 - 1}{x^2 + 1} + \frac{h}{x^2 + 1}. \quad (\text{B3})$$

For $h = 0$ we are on the line $\delta^+ = \delta^-$, which separates region 1 and 2, see Fig. 3. Furthermore, one can show that for $\tau \rightarrow \infty$ the line $\delta^+ = -\delta^-$ is given by $\delta^- = -\tanh \frac{V}{4}$ such that h is bounded in region 1 and 2 by

$$\begin{aligned} \text{region 1: } & -2 \frac{x^3 - 1}{x + 1} \leq h \leq 0, \\ \text{region 2: } & 0 \leq h \leq 2. \end{aligned} \quad (\text{B4})$$

Next, we have a look at the derivative of $\Delta_{\mathbf{i}} S_{\text{tape}}$ with respect to h , which reads:

$$\Delta_{\mathbf{i}} S'_{\text{tape}} \equiv \frac{\partial}{\partial h} \Delta_{\mathbf{i}} S_{\text{tape}} = \frac{(x^2 + x + 1) \operatorname{arctanh} \left(\frac{h-2}{x^2+1} + 1 \right) - x \operatorname{arccoth} \left(\frac{(x^2+1)(x^2+x+1)}{(h-1)x+x^4+x^3-1} \right) - (x^2 + 1) \ln(x)}{x^2 + x + 1} \quad (\text{B5})$$

and we note that $\Delta_{\mathbf{i}} S'_{\text{tape}}(h = 0) = 0$, which can be deduced by using the identities

$$\operatorname{arctanh}(z) = \frac{1}{2} \ln \frac{1+z}{1-z}, \quad \operatorname{arccoth}(z) = \frac{1}{2} \ln \frac{z+1}{z-1}. \quad (\text{B6})$$

We are done with our proof if we can show that $\Delta_{\mathbf{i}} S'_{\text{tape}} \leq 0$ for region 1 and $\Delta_{\mathbf{i}} S'_{\text{tape}} \geq 0$ for region 2 because this implies $\Delta_{\mathbf{i}} S_{\mathbf{i}} \geq 0$ for region 1 and 2. To deduce this we have a look at the second derivative, which becomes

$$\Delta_{\mathbf{i}} S''_{\text{tape}} \equiv \frac{\partial^2}{\partial h^2} \Delta_{\mathbf{i}} S_{\text{tape}} = \frac{2(x+1)(x^2+1)^2[h(1-x) + 2x(x+1)]}{(h-2)(h+2x^2)[h+2x(x^2+x+1)][hx-2(x^2+x+1)]}. \quad (\text{B7})$$

If we can show that $\Delta_{\mathbf{i}} S''_{\text{tape}} \geq 0$ for all h in region 1 and 2 and for all $x \geq 1$, we are done with the proof because this would imply that $\Delta_{\mathbf{i}} S'_{\text{tape}}$ is a monotonically increasing function and since $\Delta_{\mathbf{i}} S'_{\text{tape}}(h = 0) = 0$ we have $\Delta_{\mathbf{i}} S'_{\text{tape}} \leq 0$ in region 1 and $\Delta_{\mathbf{i}} S'_{\text{tape}} \geq 0$ in region 2 as desired. To prove $\Delta_{\mathbf{i}} S''_{\text{tape}} \geq 0$ it suffices to look at the sign of all the factors of $\Delta_{\mathbf{i}} S''_{\text{tape}}$ and to show that their overall sign is positive. The estimation of the factors can be straightforwardly done by using the bounds for h , Eq. (B4), which we will not do here. Thus, we have proven that Eq. (22) is true for $\tau \rightarrow \infty$.

Let us now turn to the case of finite τ . To begin, we have confirmed numerically that

$$\Delta N^L(\tau) = \eta \Delta N^L(\infty) \quad \text{with} \quad \eta \in [0, 1], \quad (\text{B8})$$

i.e., the number of tunneled particles becomes maximized (in absolute value) for $\tau \rightarrow \infty$. Next, for $\tau \rightarrow \infty$ Eq. (15) implies $\Delta N^L(\infty) = (\delta^+(\infty) - \delta^-)/2$ and again using Eq.

(15) for finite τ together with Eq. (B8) gives

$$\begin{aligned} \delta^+(\tau) &= \delta^- + 2\eta \frac{\delta^+(\infty) - \delta^-}{2} \\ &= (1 - \eta)\delta^- + \eta\delta^+(\infty). \end{aligned} \quad (\text{B9})$$

Using that the entropy is a concave function (i.e., $\partial_{\delta}^2 H[\delta] < 0$) yields

$$\begin{aligned} H_B[\delta^+(\tau)] &\geq (1 - \eta)H_B[\delta^-] + \eta H_B[\delta^+(\infty)] \\ &= H_B[\delta^-] + \eta \{H_B[\delta^+(\infty)] - H_B[\delta^-]\} \\ &= H_B[\delta^-] + \eta \Delta H_B(\infty), \end{aligned} \quad (\text{B10})$$

which, together with the second law in the long time limit, implies

$$\begin{aligned} H_B[\delta^+(\tau)] &\geq H_B[\delta^-] - \eta V \Delta N^L(\infty) \\ &= H_B[\delta^-] - V \Delta N^L(\tau) \end{aligned} \quad (\text{B11})$$

and, after rearrangement, we finally obtain our desired result for all τ

$$\Delta_{\mathbf{i}} S_{\text{tape}}(\tau) = V \Delta N^L(\tau) + \Delta H_B(\tau) \geq 0. \quad \text{Q.E.D.} \quad (\text{B12})$$

Appendix C: Maximum and Minimum of the Mutual Information

The first derivative of the mutual information with respect to δ_0 is given by

$$\begin{aligned} I' &\equiv \frac{\partial I(S; M')}{\partial \delta_0} \\ &= (1 - p_{\uparrow}^- - p_{\downarrow}^-) \\ &\quad \times \{ \operatorname{arctanh}(\delta_0) - \operatorname{arctanh}[h(\delta_0, \delta^-)] \} \end{aligned} \quad (C1)$$

with $h(\delta_0, \delta^-) \equiv \delta_0 - \delta_0(p_{\uparrow}^- + p_{\downarrow}^-) + \delta^-(p_{\uparrow}^- - p_{\downarrow}^-)$. Clearly, $I' = 0$ if $\delta_0 = \delta_0^{\min}$ with δ_0^{\min} from Eq. (31) and it can easily be checked that the second derivative is positive at this point, hence δ_0^{\min} truly minimizes the mutual information. We also recognize that for $\delta_0 > \delta_0^{\min}$ we have $I' > 0$ and for $\delta_0 < \delta_0^{\min}$ we have $I' < 0$ for every δ^- . Thus, the mutual information has its local maxima at $\delta_0 = \pm 1$. From Eq. (30) we obtain

$$\delta_M(\delta_0 = +1) = p_0^- + \delta^-(p_{\uparrow}^- - p_{\downarrow}^-), \quad (C2)$$

$$\delta_M(\delta_0 = -1) = -p_0^- + \delta^-(p_{\uparrow}^- - p_{\downarrow}^-). \quad (C3)$$

We now claim that $p_{\uparrow}^- \geq p_{\downarrow}^-$ in the Maxwell demon region 2 of Fig. 3. Intuitively this should be clear because a positive bias $V > 0$ favors the spin up state [see Eq. (3)] and the excess of spin up states on the tape will also preferably flip spin down states to spin up states in the system. Mathematically, however, we could only confirm this numerically. Accepting $p_{\uparrow}^- \geq p_{\downarrow}^-$ we easily

see that $|\delta_M(\delta_0 = +1)| \geq |\delta_M(\delta_0 = -1)|$, which implies that $I(\delta_0 = -1) = H[\delta_M(\delta_0 = -1)] - p_1^- H[\delta^-]$ is larger than $I(\delta_0 = +1) = H[\delta_M(\delta_0 = +1)] - p_1^- H[\delta^-]$.

Appendix D: Derivation of the effective Master Equation

We divide the evolution of the system density matrix into pieces where there is either no interaction with the bit (only the dissipative dynamics \mathcal{W} of the spin valve) or there is a sudden interaction with a bit (and no dissipation) due to the swap operation. We introduce the notation $\mathcal{J}\rho_S \equiv \operatorname{tr}_B[U(\rho_S \otimes \rho_B)U^\dagger]$ with the same U from Eq. (8) and call \mathcal{J} a ‘jump operator’. Because we are still working in the basis $\rho_S = (p_0, p_{\uparrow}, p_{\downarrow})^T$ and neglect any off-diagonal elements of the density matrix, we can represent \mathcal{J} by a matrix of the form

$$\mathcal{J} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & p_{b=\uparrow}^- & p_{b=\uparrow}^- \\ 0 & p_{b=\downarrow}^- & p_{b=\downarrow}^- \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1+\delta^-}{2} & \frac{1+\delta^-}{2} \\ 0 & \frac{1-\delta^-}{2} & \frac{1-\delta^-}{2} \end{pmatrix}. \quad (D1)$$

Note that the jump operator for the situation involving measurement and feedback [together with the choice (25)] is exactly the same, hence the ME is also the same for both cases.

The density matrix $\rho_S^{(n)}(t)$ describing the state of the system after n jumps have occurred up to time t is then given by

$$\begin{aligned} \rho_S^{(n)}(t) &= \int_{t_0}^t dt_n \int_{t_0}^{t_n} dt_{n-1} \dots \int_{t_0}^{t_2} dt_1 \gamma e^{-\gamma(t-t_n)} e^{\mathcal{W}(t-t_n)} \mathcal{J} \dots \mathcal{J} \gamma e^{-\gamma(t_1-t_0)} e^{\mathcal{W}(t_1-t_0)} \rho_S^{(0)}(t_0) \\ &= \gamma^n e^{-\gamma(t-t_0)} \int_{t_0}^t dt_n \int_{t_0}^{t_n} dt_{n-1} \dots \int_{t_0}^{t_2} dt_1 e^{\mathcal{W}(t-t_n)} \mathcal{J} \dots \mathcal{J} e^{\mathcal{W}(t_1-t_0)} \rho_S^{(0)}(t_0) \end{aligned} \quad (D2)$$

and the average state of the system can be recovered via $\rho_S(t) = \sum_{n=0}^{\infty} \rho_S^{(n)}(t)$.

Next, we take the time derivative of $\rho_S^{(n)}$, which yields three terms:

$$\frac{\partial}{\partial t} \rho_S^{(n)}(t) = -\gamma \rho_S^{(n)}(t) + \mathcal{W} \rho_S^{(n)}(t) + \gamma \mathcal{J} \rho_S^{(n-1)}(t). \quad (D3)$$

Averaging over n gives finally the ME $\dot{\rho}_S(t) = \mathcal{W}_{\text{eff}} \rho_S(t)$ with the effective Liouvillian $\mathcal{W}_{\text{eff}} = \mathcal{W}_L + \mathcal{W}_R + \mathcal{W}_B$ where $\mathcal{W}_{L,R}$ are given by Eq. (1) and $\mathcal{W}_B \equiv \gamma(\mathcal{J} - 1)$, which equals Eq. (35) of the main text.

Appendix E: Proof of Eq. (38)

The change of the Shannon entropy of the bits over an infinitesimal small time step dt is given by $dH_B(t) = H_B(t+dt) - H_B(t)$ with

$$H_B(t+dt) = - \sum_{\sigma} p_{b=\sigma}(t+dt) \ln p_{b=\sigma}(t+dt), \quad (E1)$$

$$H_B(t) = - \sum_{\sigma} p_{b=\sigma}(t) \ln p_{b=\sigma}(t) \quad (E2)$$

where $p_{b=\sigma}(t)$ denotes the probability to find the bit in state $\sigma \in \{\uparrow, \downarrow\}$. Because at every time step the old bit is replaced by a new bit initialized with the probability distribution for the incoming tape, we must set $p_{b=\uparrow}(t) =$

$(1 + \delta^-)/2$ and $p_{b=\downarrow}(t) = (1 - \delta^-)/2$. The state of the outgoing bit is then given by

$$\begin{aligned} p_{b=\uparrow}(t + dt) &= p_{b=\uparrow}(t) + \gamma dt [p_{b=\downarrow}(t)p_{s=\uparrow}(t) - p_{b=\uparrow}(t)p_{s=\downarrow}(t)] \quad (\text{E3}) \\ &= p_{b=\uparrow}(t) + dt I^B \end{aligned}$$

where we have explicitly denoted the state of the system at time t with $p_{s=\sigma}(t)$ and I^B is the current from Sec. [V A](#). Similarly, $p_{b=\downarrow}(t + dt) = p_{b=\downarrow}(t) - dt I^B$. Using these

relations and the expansion of the logarithm $\ln(1 + x) = x + \mathcal{O}(x^2)$, it is now a matter of straightforward algebra to show that

$$\begin{aligned} \frac{dH_B(t)}{dt} &\equiv \frac{H_B(t + dt) - H_B(t)}{dt} \\ &= I^B \ln \frac{1 - \delta^-}{1 + \delta^-} + \mathcal{O}(dt), \end{aligned} \quad (\text{E4})$$

which proves Eq. [\(38\)](#).